# 

[**Lecture 1: Introduction to the course. GQM model 2**](#_rb77p04hc70p)

[**Lab 1 - GQM 5**](#_narjrso7q16s)

[**Lecture 2: Hypothesis Testing 9**](#_s7oow6v9qdms)

[Lab 2 - Hypothesis Testing 10](#_ze6qm444h1vw)

[**Lecture 3 Measurement, concept, definition, fundamentals, representational theory of measurement 10**](#_avaog5kend25)

[Lab 4 - Goodness of measurement (related to Lec 3 more) 11](#_wnk4ql6ex9sh)

[**Lecture 4 - Measurement scales and functions that can be applied to scales: 13**](#_bsryubxjjorg)

[Lab 3 - Measurement scales (related to Lec 4 more) 14](#_us6ruhk4jt28)

[**Lecture 5: Introduction to Experimental Design 14**](#_drmyi2l2vvyj)

[Lab 5 - Experimental Research 17](#_zav4nwupdoye)

[**Lecture 6 - Experimental Design 18**](#_e16xc6xbfuex)

[**Easter eggs 22**](#_ijue49cd1qjt)

# 

# Lecture 1: Introduction to the course. GQM model

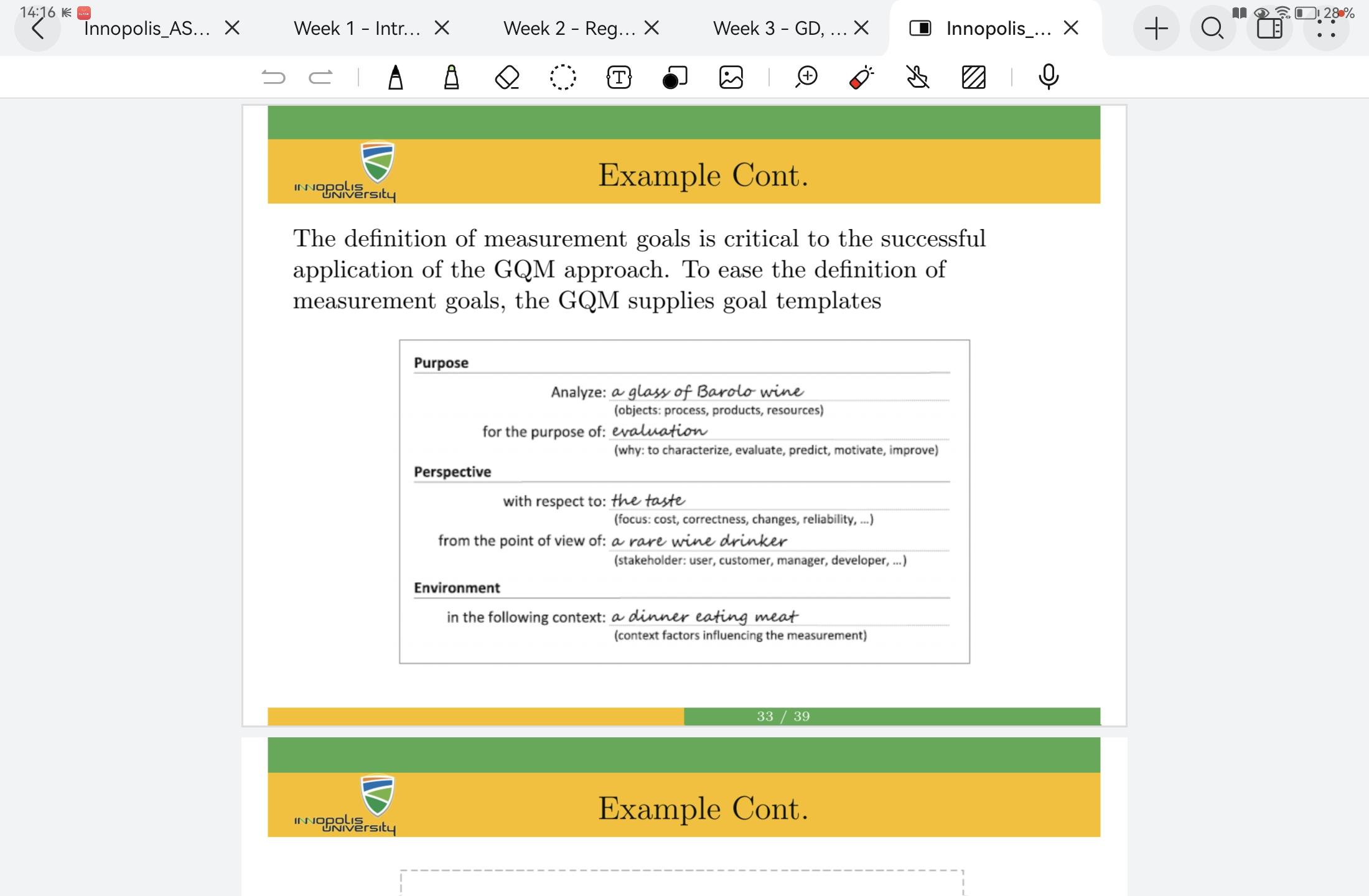
GQM is a goal oriented approach, it means that every time I want to measure something I need to link the measure to the goal.

G - goal (conceptual level)

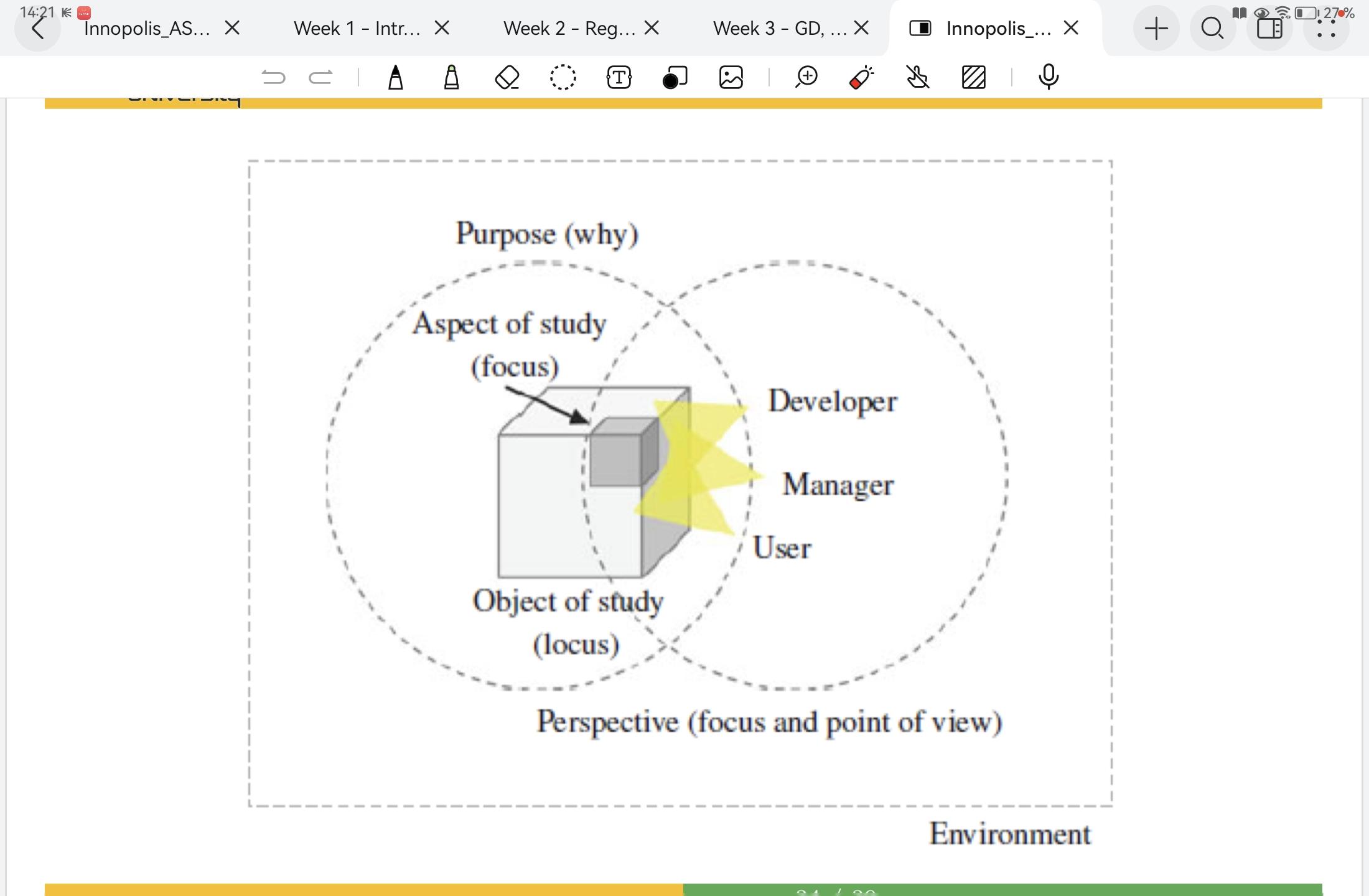
Q - question (operational level)

M - metric (quantitative level)

Goal template for GQM



GQM + Strategies



Object: the object of studies

Focus: the aspect of studies

Magnitude: the desired magnitude of improvement

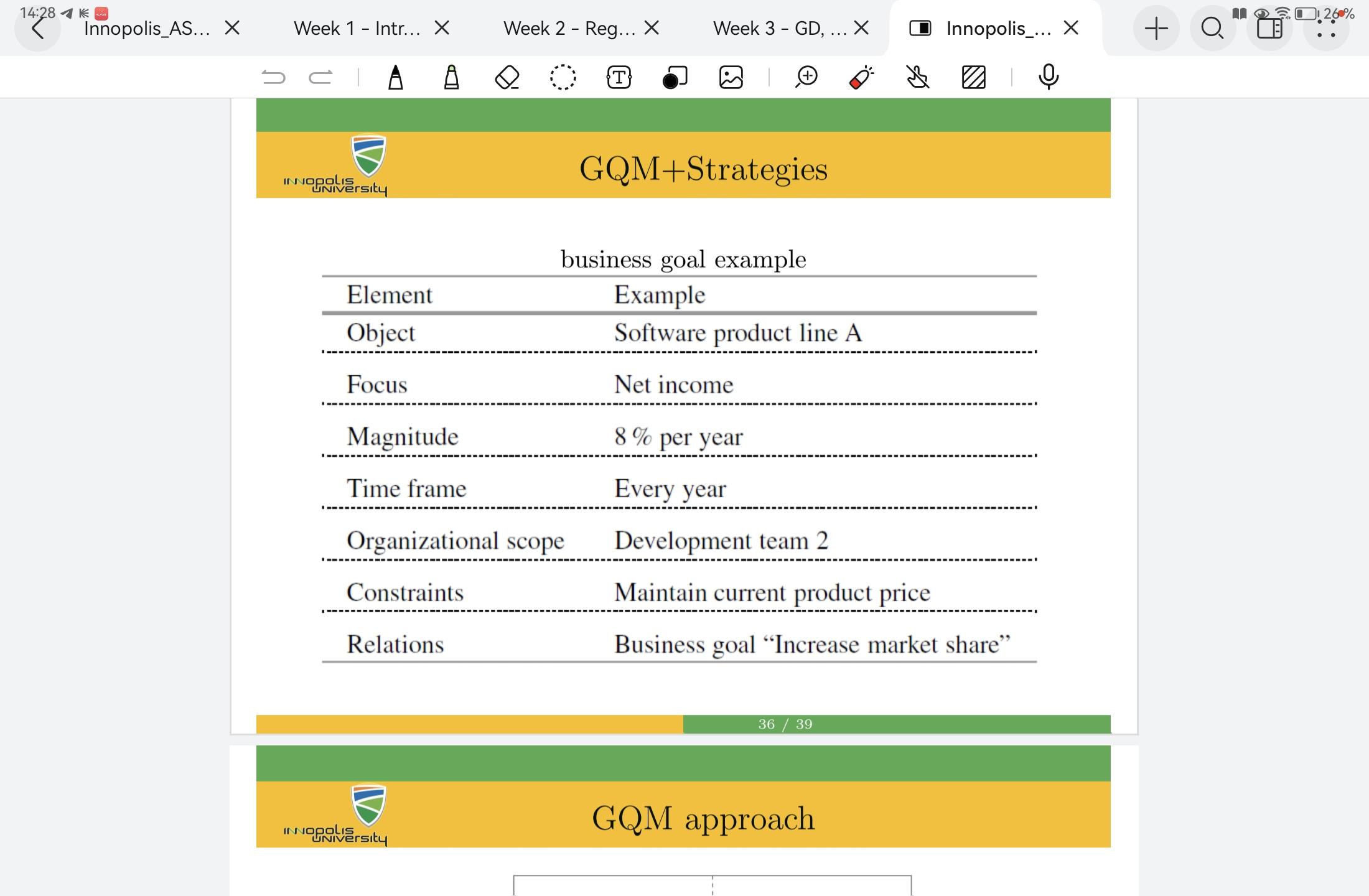
Time frame: the time frame for achieving the goal

Organizational scope: the scope of responsibility for achieving the goal

Constraints: constraints or conflicting goals

Relationships: relationships to other goals

Example



SWOT analysis

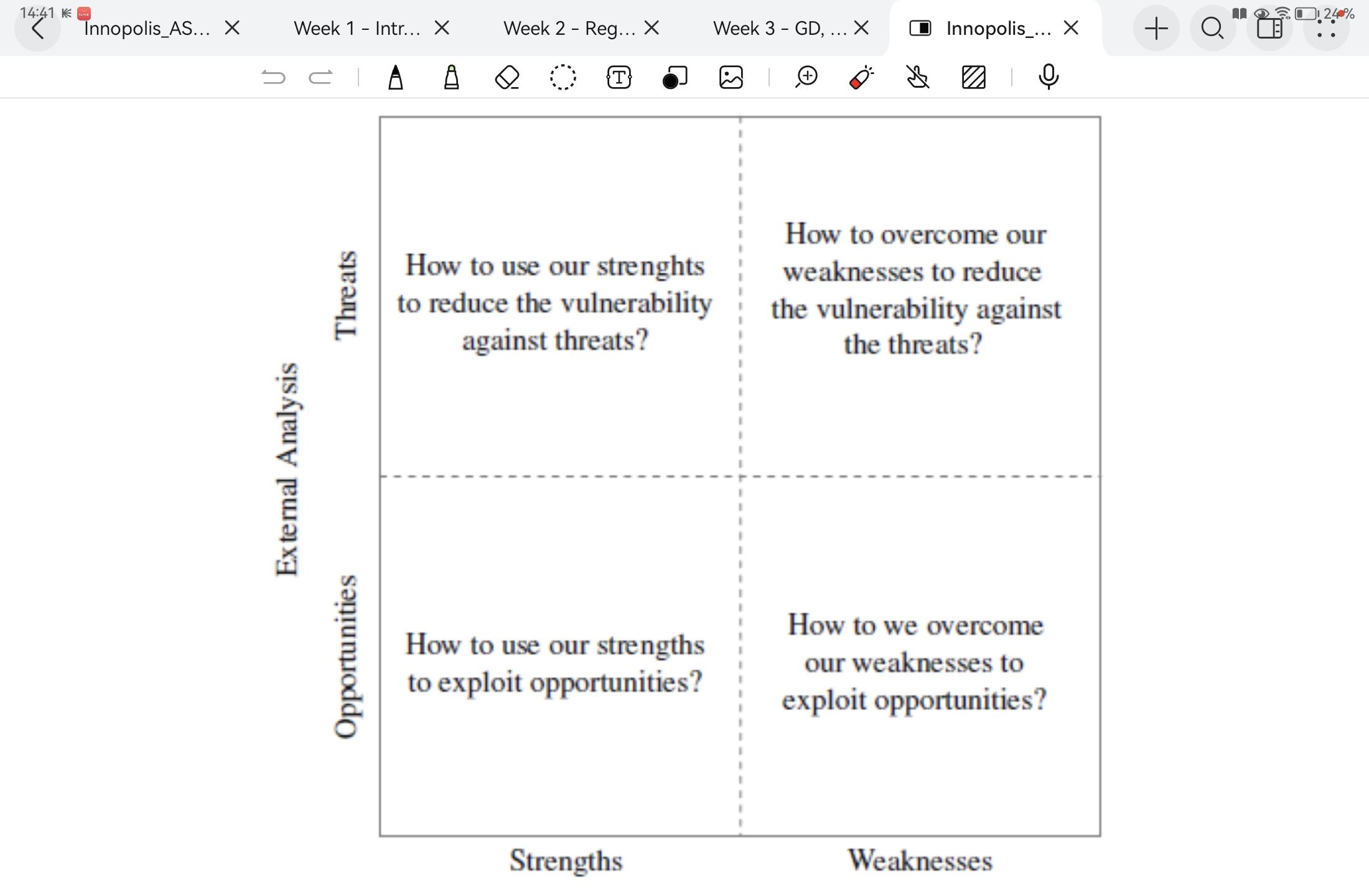
S - strengths - the organizational aspects that support the achievement of the objective

W - weaknesses - the organizational aspects that inhibit the achievement of the objective

O - opportunities - the external aspects that support the achievement of the objective

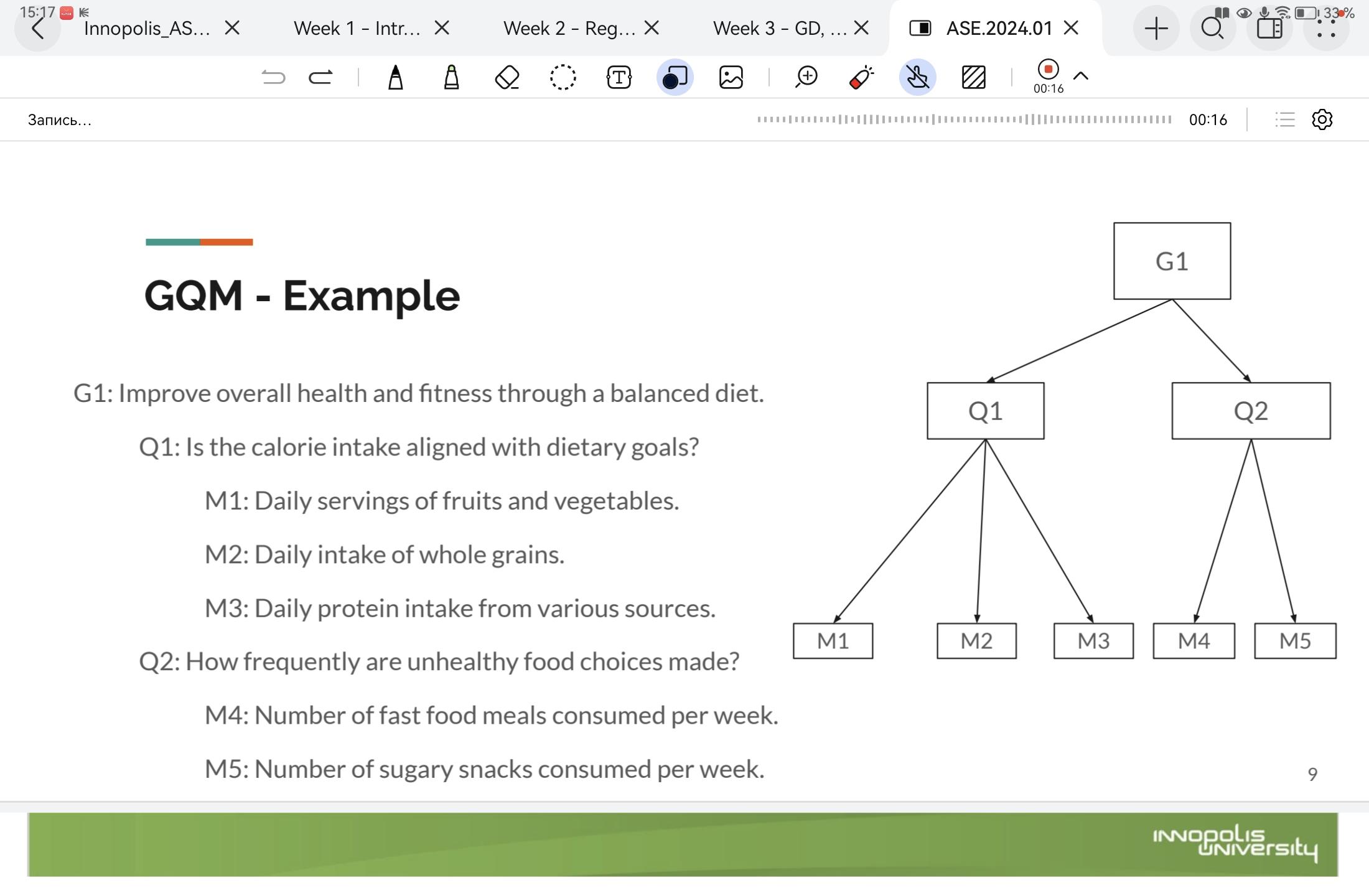
T - threats - the external aspects that inhibit the achievement of the objective

Some questions to think:

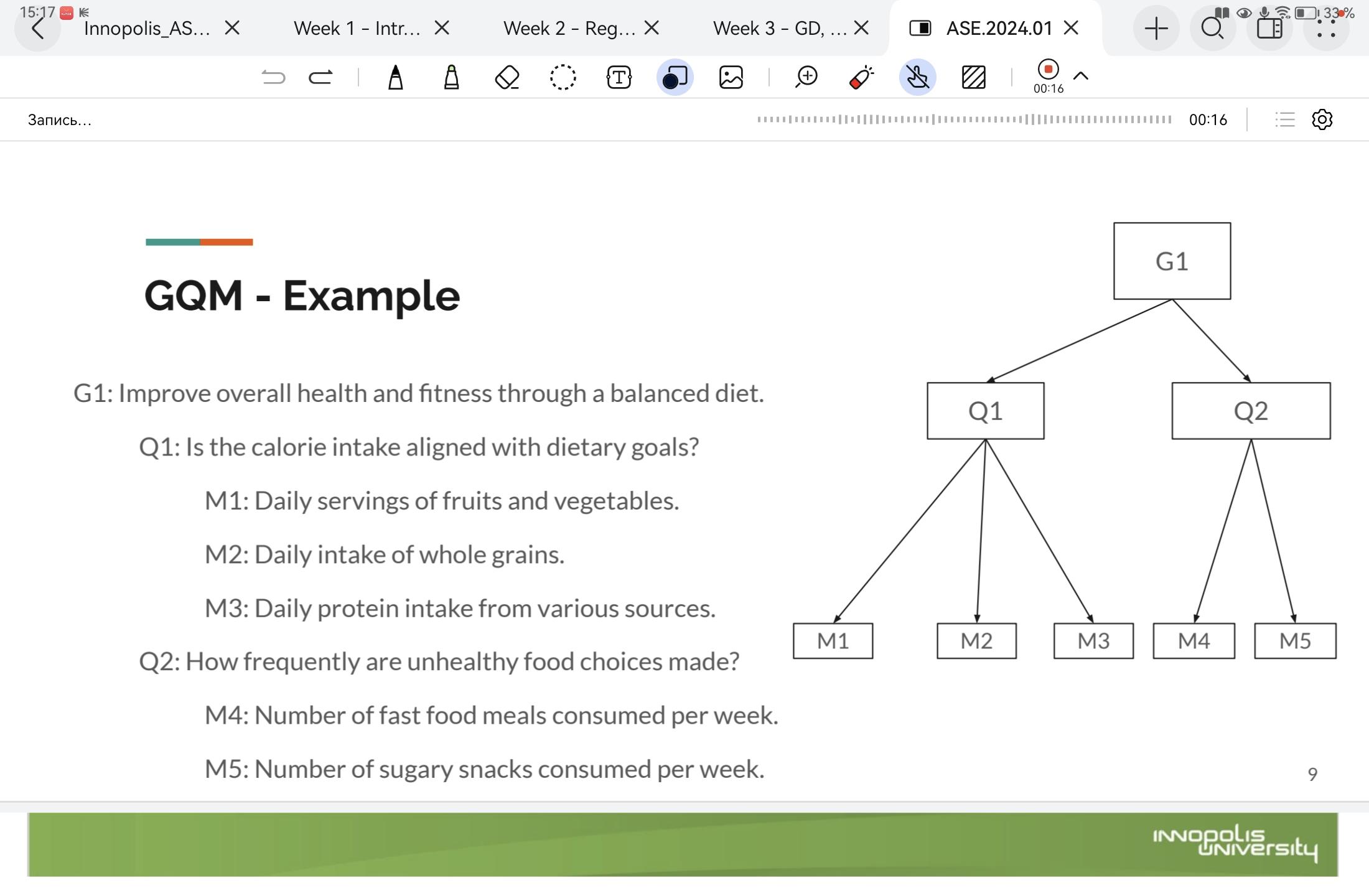


## **Lab 1 - GQM**

Example of GQM:



GQM paradigm



Goal definition template

**Analyze** {the name of activity or attribute}

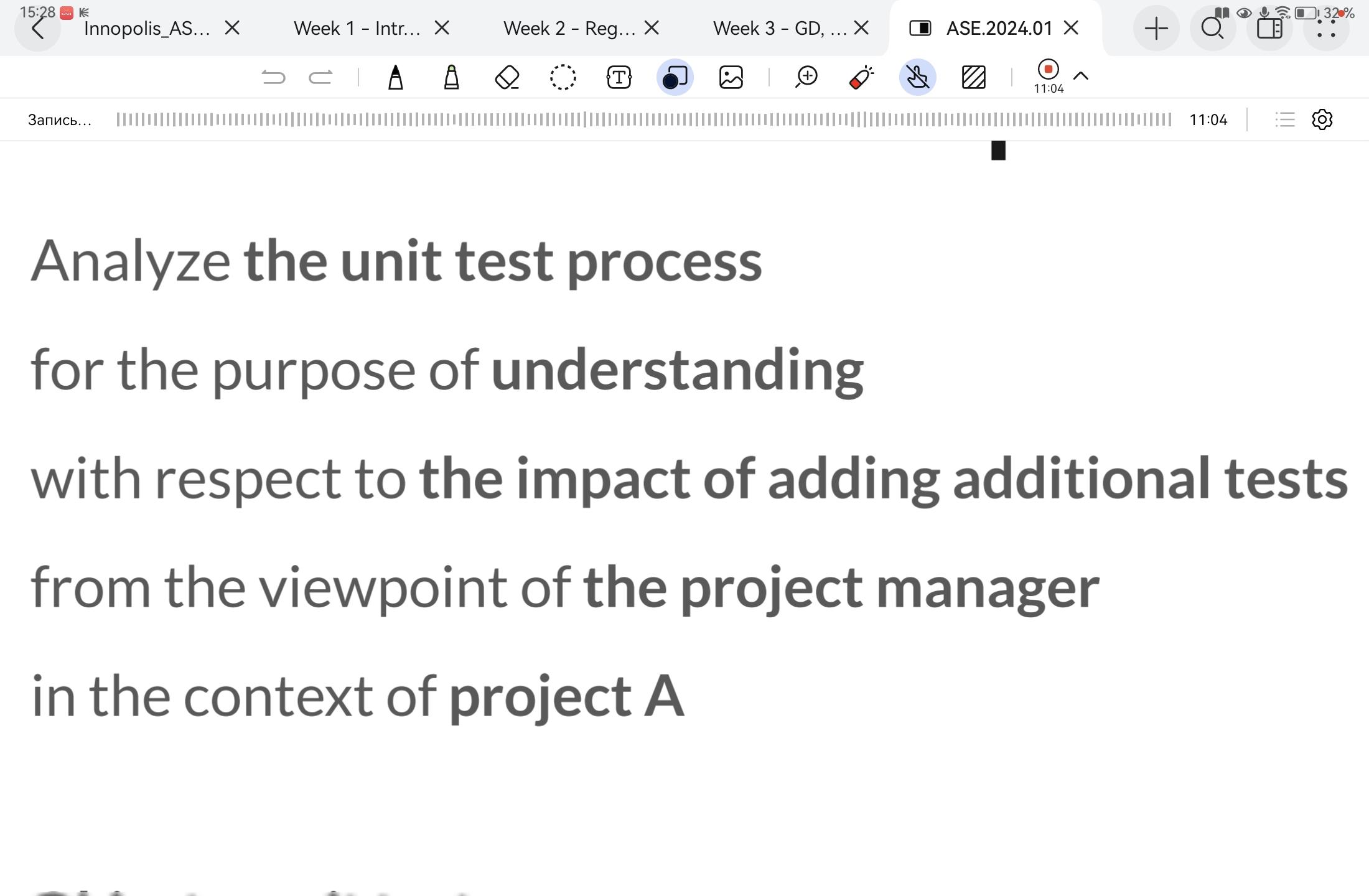
**for** **the** **purpose of** {overall goal}

**with respect to** {the aspect to be considered}

**from the view point of** {interested people}

**in the context of** {environment}

Example



SMART

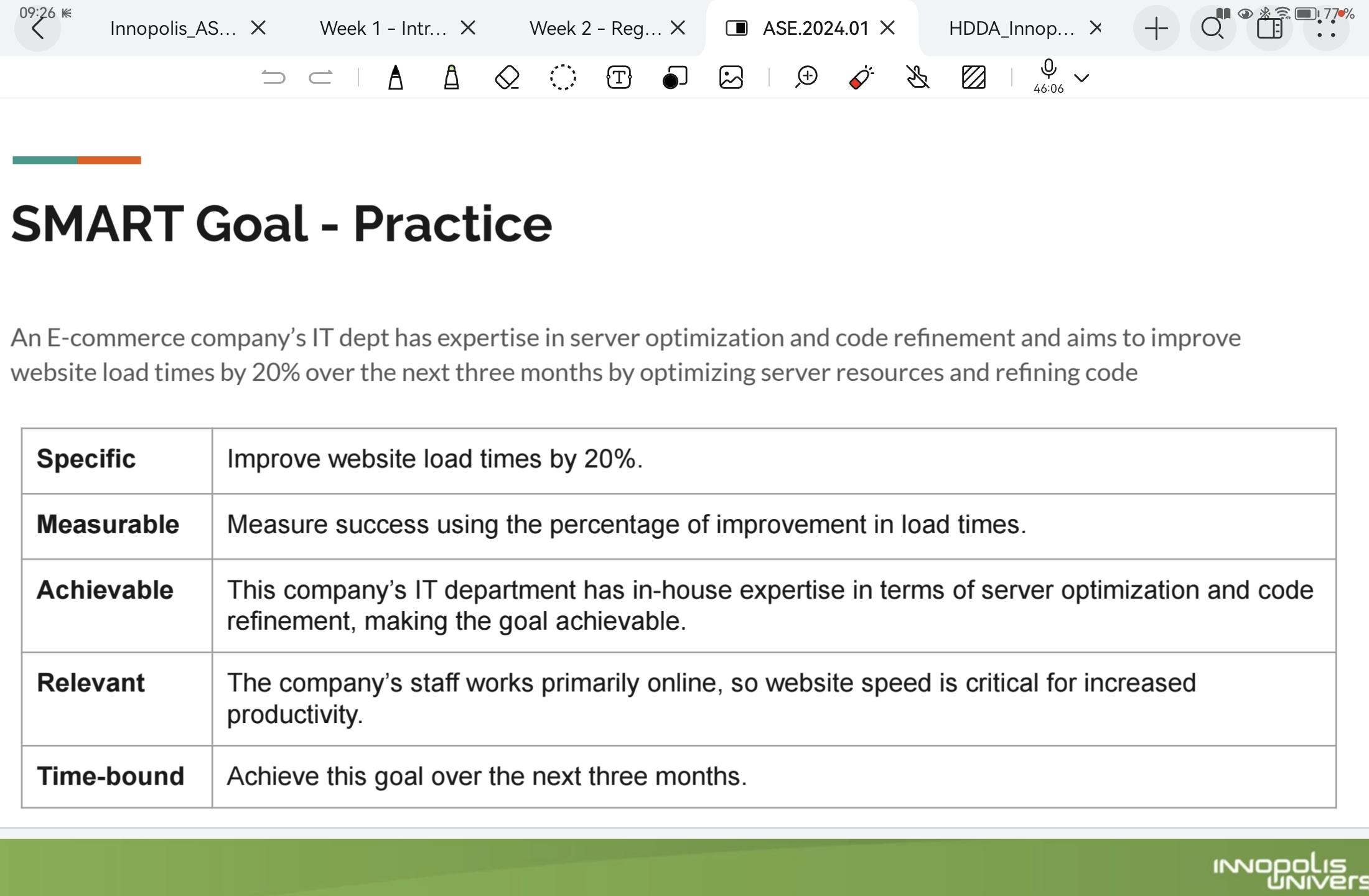
S - specific - provide a clear description of what we need to achieve

M - measurable - include a metric with a target that indicates success

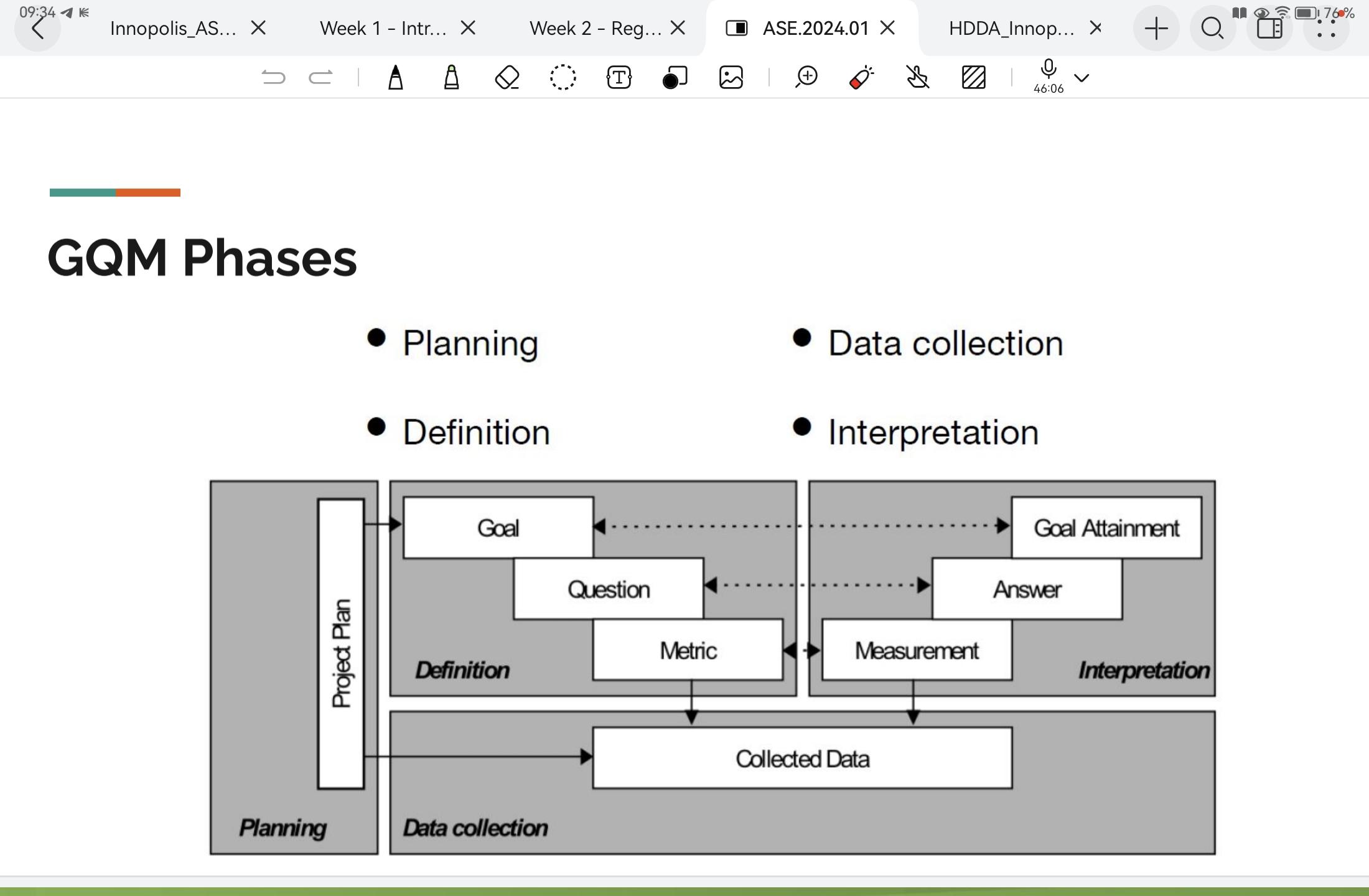
A - achievable - set a challenging target, but keep it realistic

R - relevant - keep your goal consistent with higher-level goals

T - time-bound - set a date for when your goal needs to be achieved



GQM phases



# Lecture 2: Hypothesis Testing

**Definitions:**

**Descriptive statistics** - *compute* properties (*statistics*) of the sample to present the data in a meaningful way.

**Inferential statistics** - analyze samples to make predictions about overall population (*estimates* the population *parameters* from the sample statistics) (do generalization about the population).

**Population** – set of objects that are studied in a task. It could be all software engineers in the world, employees of XXX company, MSIT students, etc.

**Sample** – finite set of objects from the population. *Representative samples* - samples that represent main properties of the population.

**Conditions of “good” estimators:**

* **Unbiased** - when the center of the sampling distribution for the estimate is the same as that of the population.
* **Consistent** - the value of the estimator (of sample) converges to the value of the parameter (of population) as the sample size increases.
* **Relatively Efficient** - the best estimator should have the smallest variance when compared to other estimators.

**Types of estimates:**

* **Point estimates** - the single best guess about the value of parameter
* **Interval estimates (confidence interval)** - the interval that contain the true value (of the corresponding parameter) with the specified probability

**Hypothesis testing** - is a systematic procedure for deciding whether the results of a research study support a particular theory which applies to a population (*uses a sample to evaluate a hypothesis about a population*).  
https://lstat.kuleuven.be/training/coursedescriptions/Goodyear/critical\_region.pdf

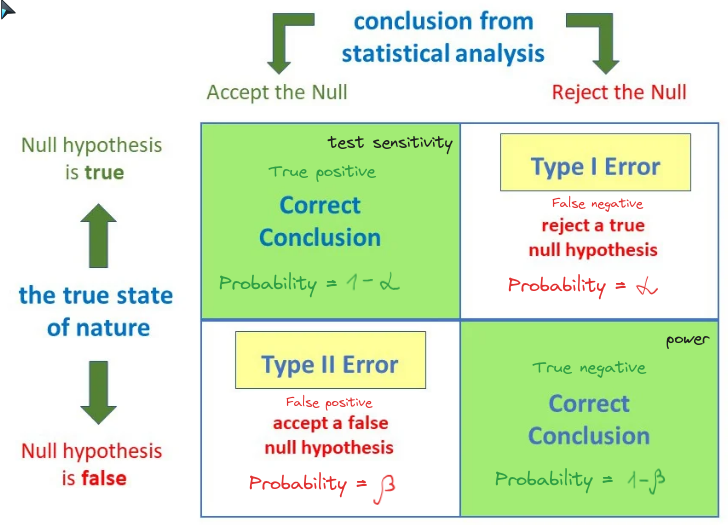
* **Null hypothesis** - expresses that there is no relationship between variables/no differences between groups/we don’t see an effect of smt.
* **Alternative hypothesis** - otherwise, hypothesis of a specific effect.

*If the null hypothesis is rejected, then the alt hypothesis can be accepted. If the alt hypothesis is rejected, then the null hypothesis is failed to reject.*

* **Confidenceinterval** (acceptance region) - is a set of values for the test statistic for which the null hypothesis is failed to reject. *i.e. if the observed test statistic is in the confidence interval then we reject the alt hypothesis.*
* **Critical region** is a set of values for the test statistic for which the null hypothesis is rejected. *i.e. if the observed test statistic is in the critical region then we reject the null hypothesis and accept the alternative hypothesis.*
* **P-value** - is a probability describing how likely it is that our data would have occurred by *random chance* (i.e., that the null hypothesis is true). *The smaller the p-value, the less likely the results occurred by random chance, and the stronger the evidence that we should reject the null hypothesis. (the probability that our results are biased by noise (erroneous))*

**Steps of hypothesis testing:**

* Specify H0 and H1
* Determine the appropriate test statistic (mean, var, std)
* Determine the critical region (by level of significance 𝛂)
* Compute the value of the test statistic
* Make decision



**Type I** (𝛂)*(significance level/size of the critical region)* - when the null hypothesis is rejected, which is actually true

**Type II** (𝛃)*(blindness)* - when the null hypothesis is failed to reject, which is actually false

## **Lab 2 - Hypothesis Testing**

# **Lecture 3 Measurement, concept, definition, fundamentals, representational theory of measurement**

**Definitions:**

**Measurement -** is the process by which numbers or symbols are assigned to attributes of entities in the real world in such a way as to describe them according to clearly defined rules. *(is a map/transformation from the empirical world to a world of numbers, then we can perform operational numbers, and then go back to the real world and determine the properties of the real world.)*

**Measure -** is a product of attaching a numerical value (or symbol) to an attribute of an entity for the purpose of to study the relationship between entities that exist.

**Empirical Relations** - the data we obtain as measures should represent attributes of the entities we observe, and manipulation of the data should preserve relationships that we observe among the entities

**Representational condition -** is what ensures that the relationships which are true in the empirical world are also true in the mathematical world.

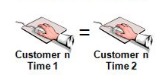
**Types of measures:**

* **By by point of view:**
* **Objective measure** - when no matter the device we use to take, the result remains the same
* **Subjective measure** - when the result depends on the instrument (or person) we use to take, and changing the measurement instrument can change the result
* **By calculating:**
  + **Direct measure -** which is taken directly from the instrument
  + **Indirect measure -** which is combined from the direct measures

**Key stages of formal measurement:**

1. Identify attribute for some real-world entities
2. Identify empirical relations for attribute
3. Identify numerical relations corresponding to each empirical relation
4. Define mapping from real world entities to numbers (do measuring)
5. Check that numerical relations preserve and are preserved by empirical relations (representative condition)

## **Lab 4 - Goodness of measurement (related to Lec 3 more)**

**Criterias of assessing measurement quality:  
(https://www.researchgate.net/profile/Ram-Bajpai-3/publication/271186978\_Goodness\_of\_Measurement\_Reliability\_and\_Validity/links/5503164b0cf24cee39fd591b/Goodness-of-Measurement-Reliability-and-Validity.pdf)**

* ***Sensitivity*** *(property of the instrument) – accuracy of the measuring instrument.*
* **Reliability** - the degree to which measures are free from random error and consequently give consistent and stable results:
  + **Stability(repeatability)** - preserves the same results over the time in the measurement of a concept
    - **Test-retest & intra-rater** - the reliability coefficient obtained by repetition of the same measure on a second time.
    - **Parallel form** - when responses on two comparable sets of measures using the same construct are highly correlated 
  + **Consistency(homogeneity)** -
    - **Inter-rater** - test of consistency respondents’ answers to all concepts, they’ll be correlated with one another
    - **Internal consistency** - indicates the uniformity of the items and items should be capable of independently measuring the same concept
* **Validity** - a test of how well an instrument that is developed measures the particular concept it is intended to measure:
  + **Content** - ensures that the measure includes an adequate and representative set of all items that tap the concept.
  + **Face Validity** - assessed by initial logic and by stakeholders
  + **Criterion** - is used to predict future or current performance whether it correlates test results with another criterion of interest.
    - **Predictive** - indicates the ability of the measuring instrument to differentiate among individuals with reference to a future criterion.
    - **Concurrent** - offers a way of establishing a test’s validity by comparing it to another similar test that is known to be valid
  + **Construct** - testifies to how well the results obtained from the use of the measure fit the theories around which the test is designed:
    - **Convergent** - is established when the scores obtained with two different instruments measuring the same concept are highly correlated;
    - **Discriminant(Divergent)** - is established when, based on theory, two variables are predicted to be uncorrelated, and the scores obtained by measuring them are indeed empirically found to be so.

**Measurement errors:**

* **Random error** - varies unpredictably from one measurement to another and is unavoidable, but accumulates around the true value
  + Reducing: take repeated measurements, increase sample size, increase the sensitivity of instrument, control other variables.
* **Systematic error** - has the same value for every measurement and can be avoided by calibrating equipment, but if left uncorrected, can lead to measurements far from the true value.
  + Reducing: triangulation (use multiple techniques/methods/instruments to record observations), regular calibration (comparing with known standard value)

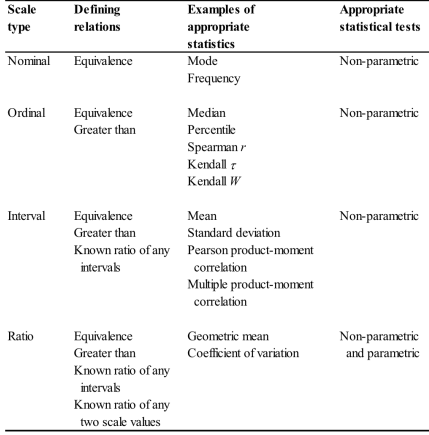
# **Lecture 4 - Measurement scales and functions that can be applied to scales:**

**Definitions:**

**Measurement scale** - is a class of mapping that links empirical and number relations with specific properties. *Scales are important because they tell us what kind of statistical analysis (type of operations) we can do with the data.*

**Scales:**

1. **Nominal** - categories, which aren't compared to each other (calc. = mode, *count, equality*)
2. **Ordinal** - categories can be putted into meaningful arrangement (distances between categories aren't necessarily equal) (calc. += median, *comparison, range*)
3. **Interval (*Y = aX+b*)** - equal distances between points (scores), arbitrary (not absolute) zero (can't use ratios) (calc. += mean, variance, std, *addition, subtraction*)
4. **Ratio (*Y=aX*)** - can use rations, there is absolute zero (calc. += *multiplication, division*)
   1. **Absolute (*Y=X*)** - just are *‘counts*’, unique in the nature



* **Meaningful measures** - if their truth value does not change when the measure is subject to transformation, i.e. they are defined on the appropriate scale.

## **Lab 3 - Measurement scales (related to Lec 4 more)**

Many examples… and nothing else is new

# **Lecture 5: Introduction to Experimental Design**

**Definitions:**

**Experiment** is a test under controlled conditions that is made to demonstrate a known truth or examine the validity of a hypothesis.

**Treatment** - the exposure of a group to an experimental variable or event, the effects of which are to be measured.

**Subjects** - experimental units, who are exposed to treatments.

**Observation** - the process of measuring the dependent variable within the experiment. Experimenter performs observation(s) before and/or after the treatment intervention.

**Experimental group -** is the group that receives the treatments being tested in an experiment.

**Control group -** is the group in an experiment that does not receive the treatments we are testing.

**Experimental design/research** systematically manipulates one or more variables in order to evaluate how this manipulation impacts an outcome of interest. An experiment isolates the effects of this manipulation by holding all other variables constant.

**Types of sample:**

**Random sample** - each item in the population is selected randomly, i.e., informally has an equal probability of being selected.

**Convenience sample** - items are chosen on their convenience and availability (only few persons agreed to participate in our experiment)

**Characteristics of experiment**

- **Random assignment** - is the process of assigning individuals at random to groups or to different groups in an experiment. (*Random assignment != Random selection*) (**Random selection** - is the process of selecting a sample from a population, so that the sample is representative of the population and you can generalize results obtained during the study to the population).

- **Control over extraneous factors:**

- (**Extraneous factors)** - are any influences in the selection of the participants, the procedures, the statistics, or the design likely to affect the outcome and provide an alternative explanation for our results than what we expected.

- **Pretest** - provide a measure on some attribute or characteristic that you assess for participants in an experiment before they receive a treatment.

- **Posttest** - is a measure on some attribute or characteristic that is assessed for participants in an experiment after a treatment.

- **Covariates** - are variables that the researcher controls for using statistics and that *relate to the dependent* variable but that don't relate to the independent variable.

- **Matching** - is the process of identifying one or more characteristics that *influence the outcome* and assigning items with that characteristics *equally* to the experimental and control groups.

- **Homogeneous sampling** is selecting for experimental and control groups items which vary little in their characteristics.

- **Blocking variable** - is a variable the researcher controls before the experiment starts by dividing (or “blocking”) the items into subgroups (or categories) and analyzing the impact of each subgroup on the outcome.

- **Manipulation of the treatment conditions**

- **Treatment variables -** are independent variables that the researcher manipulates to determine their effect on the dependent variable.

- **Outcome measures**

- **Group comparisons**

- **Threats to validity -** experiments might lead to wrong conclusion, meaning that the results would not be “valid”

- **Internal validity** - a measure that ensures the results and trends seen in an experiment are actually caused by the treatment and not some other factors underlying the process (*communication across the teams*).

- **History** - any specific event (unexpected) that occurs while the experiment is in progress.

- **Regression to the mean** - the natural tendency for extreme scores to regress or move towards the mean.

- **Mortality** - if groups lost participants (e.g., due to dropping out of the experiment) they may not be equivalent.

- **Testing** - a pretest may confound the influence of the experimental treatment; using a control group mitigates this concern

- **Maturation** - normal changes over time (by natural) (not specific to the particular events) (e.g., fatigue, aging or improving skill) might affect the dependent variable; using a control group mitigates this concern.

- **Selection** - if randomization is not used to assign participants, the groups may not be equivalent

- **Instrumentation** - changes in the calibration of a measuring instrument or changes in the observers or scorers used may produce changes in the obtained measurements

- **Diffusion of treatments** - when the experimental and control groups can communicate with each other, the control group may learn from the experimental group information about the treatment.

- **Compensatory rivalry** - if you publicly announce assignments to the control and experimental groups, compensatory rivalry may develop between the groups because the control group feels that it is the “underdog”

- **Interactions with selection** - a bias in selection may produce subjects that are more or less sensitive to the experimental treatment.

- **Compensatory equalization** - when only the experimental group receives a treatment, an inequality exists that may threaten the validity of the study. The benefits of the experimental treatment need to be equally distributed among the groups in the study

- **Resentful demoralization** - when a control group is used, individuals in this group may become resentful and demoralized because they perceive that they receive a less desirable treatment than other groups.

- **External validity(generalizability)** - a measure that shows the validity of the extent to which the results of a study can generalize to other situations, people, settings, treatment variables, and measures.

- **Reactive effect of testing(Interaction of testing and treatment)** - pretest might increase or decrease the respondent’s sensitivity or responsiveness to the experimental variable.

- **Interaction of selection and treatment** - the inability to generalize beyond the groups in the experiment. *(Factors like the setting, time of day, location, researchers’ characteristics, etc. limit generalizability of the findings.)*

- **Reactive effects of experimental arrangements** - inability to generalize from the *setting* where the experiment occurred to another *setting*.

- **Multiple-treatment interference** - occurs when treatments are applied to the same respondents, but the effects of prior treatments are not usually erasable

- **Interaction of history and treatment** - when the researcher tries to generalize findings to past and future situations

- **Construct validity** - is about how well a test measures the concept it was *designed* (*theoretical meaning of a concept*) to evaluate. (whether we apply the “right” analysis, e.g., if we apply the mean on an ordinal scale, it isn’t proper) / whether a scale or test measures the construct adequately).

- **Convergent validity** tests that confirm that are expected to be related are, in fact, related.

- **Divergent validity** tests that confirm that should have no relationship do, in fact, not have any relationship.

## **Lab 5 - Experimental Research**

**Independent variable (IV)** - what researcher manipulates

**Dependent variable (DV)** - what change due to the change in IV

**Confounder** - is a variable that influences both the DV and the IV, causing a false association.

**Covariate** - is a variable that influences only the DV.

**Moderator** - is a variable that affects the strength and direction of the relationship between DV and IV.

**Mediator** (“middle man”) - is a variable that explains how the IV and DV are related.

**Steps in conducting experimental research:**

* Decide if an Experiment addresses your research problem
* Form hypotheses to test cause-and-effect relationships
* Select an experimental unit and identify study participants
* Select an experimental treatment
* Choose a type of experimental design
* Conduct the experiment
* Organize and analyze the data
* Develop an experimental research report

# **Lecture 6 - Experimental Design**

**Steps in conducting experimental design:**

1. Define the overall GQM and metrics

2. Decide if an Experiment addresses your research problem

3. Formulate Hypothesis H0 and H1

4. Select a Population and a Sample

5. Select an Experimental Treatment and introduce it

6. Choose a type of the Experimental Design

7. Conduct the Experiment

8. Organize and analyze the data

**Experiment Design Structure - Schema:**

**Research variables:**

1. **Independent variable IV** (the intervention)

2. **Dependent variable** **DV** (the outcome)

3. **Extraneous variables EV** (other, potentially confounding factors)

**Design elements:**

1. **Manipulation -** the ability to influence or direct the **IV**

2. **Control** - the ability to direct or influence important **EV** and study measurements

3. **Randomization -** unbiased [random] subject assignment to each group

**Designs:**:

* **Prospective** - the events of interest haven't yet occurred when the study begins
* **Retrospective** - the events of interest all occurred before the onset of the study

**Taxonomy of Experimental Designs (categorization):**

**Pre-experimental designs (non-experimental):**

Have one or none of the core experiment design elements

Lack: manipulation, randomization, control (often)

Are generally *retrospective*

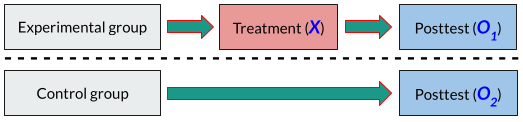
Have the *lowest* scientific validity

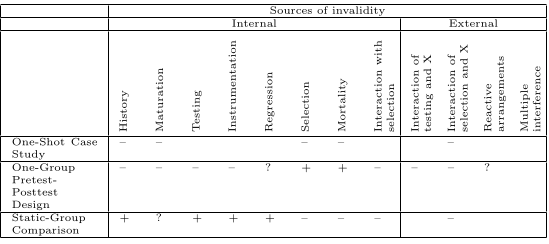
1. **The One-Shot Case Study** design - compare the post-test results to the expected results. Potential threats to internal validity: History, Maturation, Selection, Mortality.



1. **One-Group Pretest-Posttest** design - compare the post-test results to the pre-test results. Have a pretest, but don't have any control group. Potential threats to internal validity: History, Maturation, Testing, Instrumentation



1. **Static-Group Comparison** design(quasi-experimental) - the difference between the two groups is the result of the experiment. Possible threats to internal validity: selection, mortality.  
   



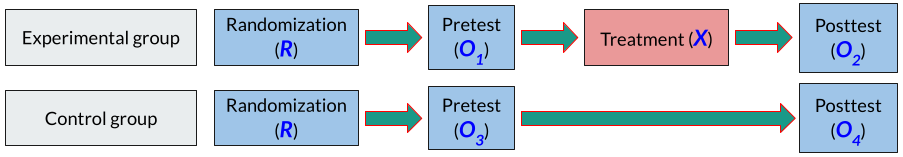
**True experimental designs**

Have all 3 design elements

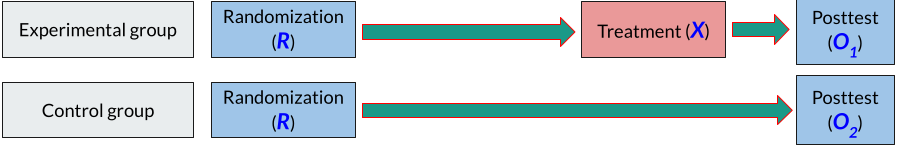
Are always prospective

Have high scientific validity

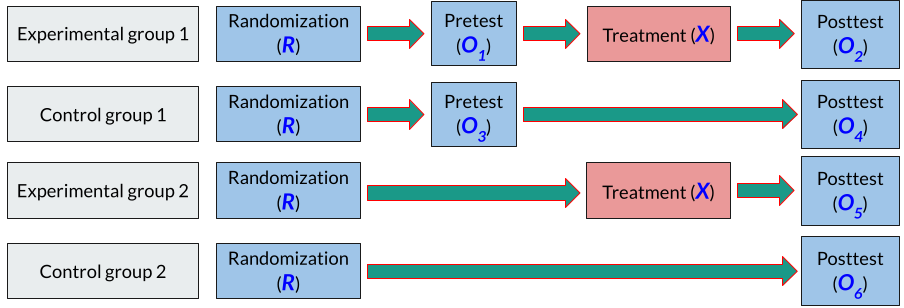
1. **Pretest-Posttest Control Group Design** - equivalence of groups is achieved by randomization.

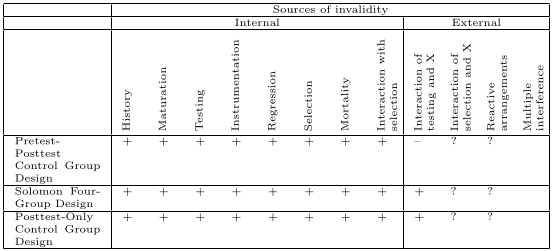


1. **Posttest-Only Control Group design** - this design controls for any confounding effects of a pretest and is a popular experimental design. The participants are randomly assigned to groups, a treatment is given only to the experimental group, and both groups are measured on the posttest.

****

1. **Solomon Four-Group -** This design deservedly has higher prestige and represents the first explicit consideration of external validity factors.  
   A special case of a 2 × 2 factorial design, this procedure involves the random assignment of participants to four groups. Pretests and treatments are varied for the four groups. All groups receive a posttest.





**Quasi-experimental designs**

Only have one or two design elements

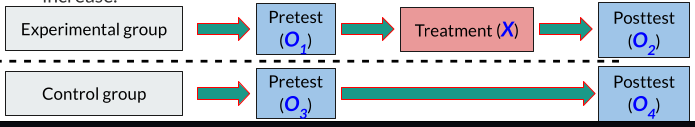
Have *manipulation* or *control*

Generally lack randomization

Are generally *prospective* in nature

Are moderate in *scientific validity*

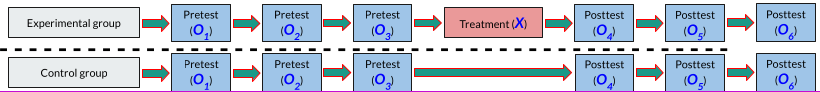
1. **Nonequivalent Control Group -**

****

1. **(Interrupted) Time-series design.** Internal threats: history, instrumentation (maybe), external: interaction of X&testing, interaction of X&selection(maybe).

****

1. **Multiple time series design.** Internal: interaction of selection & history (could be a threat). External: interaction of X&testing, interaction of X&selection.

****

1. **Regression discontinuity design -**

**Factorial Designs -** In some experimental situations, it is not enough to know the effect of a single treatment on an outcome; several treatments may, in fact, provide a better explanation for the outcome.

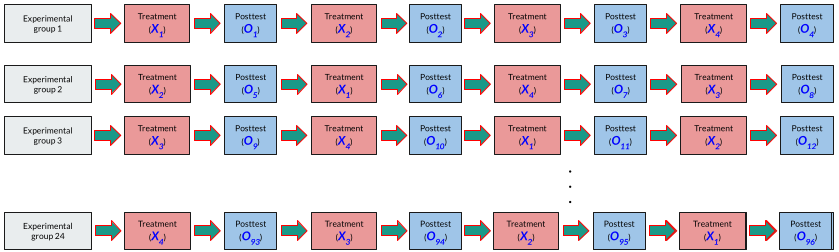
**Main effects -** the influence of each independent variable on the outcome in an experiment.

**Interaction effects** - when the influence on one independent variable depends on (or co-varies with) the other independent variable in an experiment.

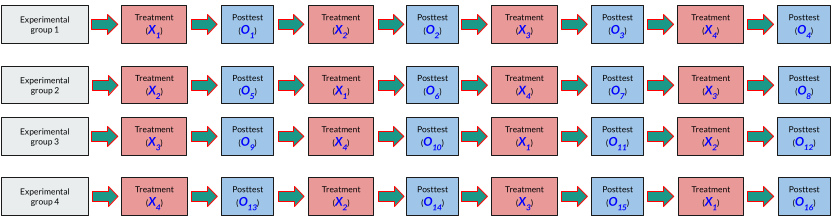
1. **Within Subject Designs**

****

1. **Counterbalanced design (n!xn)**

****

1. **Latin square design**

****

# Easter eggs

**1 var**

1. Explaining the meaning of the following terms in the context of research design:
   1. Extraneous variables - are any influences in the selection of the participants, the procedures, the statistics, or the design likely to affect the outcome and provide an alternative explanation for our results than what we expected.
   2. Treatment - the exposure of a group to an experimental variable or event, the effects of which are to be measured.
2. Fill the gaps with the correct answers. Note that each answer may contain more than one word, and some gaps may have more than one correct answer, but you need to write only one correct answer in each gap:
   1. A pre-experimental design which includes a control group is *static group comparison*
   2. Interaction of testing and treatment *is one of the external* threats which is not controlled in pretest-posttest control group design
   3. Analyzing the main effects in factorial designs are useful to understand the influence of each independent variable on the outcome
   4. Reliability is the degree to which measures are free from random error.
3. Imagine a study in which you will visually present participants with a list of 20 words, one at a time, wait for a short time, and then ask them to recall as many of the words as they can. Two groups are randomly assigned. The group that is in the stressed condition, they are told that they might also be chosen to give a short speech in front of a small audience, while the other group which is in the unstressed condition, they are not told that they might have to give a speech.
   1. Identify the following experimental components if they exist:
      1. experimental unit, dependent variable, independent variable, experimental group, control group:
         1. Experimental unit - participants of study
         2. Dependent variable - count of recalled words
         3. Independent variable - stressed condition in a short speech in front of a small audience
         4. Experimental group - a group of participants that are exposed to a stress condition
         5. Control group - a group of participants which is in the unstressed condition
   2. Suggest a specific research design for this study which can avoid history threat and specify two possible threats to its validity. Justify your answer:  
      Posttest only Control Group Design (because it is randomly assigned and don’t have pretest) can avoid *history* threat because of comparison to control group, but ‘interaction of selection and treatment’ and ‘reactive effects of experimental arrangements’ can become the threats to external validity of this design since it is пиздец.
4. Cognitive load researchers have used varying subjective techniques to quantify the cognitive load of developers. In cognitive psychology, cognitive load refers to the amount of working memory resources used.  
   Given 10 tasks from the sprint backlog in an Agile software development, two senior developers are asked to estimate the cognitive load required by tasks assuming that each task should be performed by a junior developer. For each task, they set a label (A: “the task requires zero help from someone else”. B: “I am not sure if the task can be done even with help”, C: “it requires some assistance with steps of the task”, D: “s/he will give up, and the task may cause a headache”). They also estimate the number of hours to complete the task. The data are given as follows:...
   1. State one goal appropriate for this research following the goal definition template of GQM.
   2. Two metrics are used to estimate the cognitive load required by tasks. Identify them and specify their measurement scales? Justify your answer.
   3. Calculate Cohen’s kappa coef
   4. In the 3rd sprint …
      1. Formulate your hypothesis
      2. …

**2 var**

1. Explain the meaning of the following terms in the context of research design
   1. Regression to the mean - the natural tendency for extreme scores to regress or move towards the mean.
   2. Experimental group - is the group that receives the treatments being tested in an experiment.
2. Fill the gaps with the correct answers. Note that each answer may contain more than one word, and some gaps may have more than one correct answer, but you need to write only one correct answer in each gap:
   1. A pre-experimental design which includes a pretest without a control group is One-Group Pretest-Posttest
   2. History is one of the internal threats which is not controlled in (interrupted) time series design.
   3. Analyzing the interaction effects in factorial designs are useful to understand when the impact on one independent variable depends on the other independent variable in an experiment.
   4. Type II error in hypothesis testing is the failure to reject the null hypothesis when it is actually false.
3. … the same as var 1.
4. … the same as var 1.